Sloan School of Management
Massachusetts Institute of Technology

15.013 – Industrial Economics for Strategic Decisions
Professor Robert S. Pindyck

Lecture Notes
on
Vertical Structure

(June 2009)

These notes cover a number of topics related to the vertical structure of markets. I will begin by reviewing the problem of *double marginalization*, which occurs when one or more firms selling to each other along a vertical chain have market power. You were introduced to double marginalization in 15.010; my objective here is simply to review the concept, and discuss strategies that firms sometimes use to deal with the problem.

Next, I turn to the practice of *monopsonistic price discrimination*. Often, buyers of intermediate inputs have monopsony power. The simplest way of exercising monopsony power is by reducing the quantity purchased. (See Sections 10.5 and 10.6 of Pindyck and Rubinfeld, *Microeconomics*, for a review of monopsony power.) However, sometimes firms can utilize *price discrimination* as a means of exploiting their monopsony power. We will see how this can be done, focusing in particular on the purchase of timber by paper companies.

Then, I will discuss issues that arise when a firm depends on *downstream distributers* to distribute and sell its products. Examples are bottlers in the case of soft drinks, and dealerships in the case of automobiles. What restrictions, if any, should be imposed on downstream distributers to maximize the benefits to the upstream manufacturers?

Lastly, I discuss *franchising*, a form of vertical structure that has become very common in many service-related industries. As we will see, franchising can be an effective way of dealing with problems of asymmetric information and incentive design.

1 Market Power and Double Marginalization

Often, one or more firms selling to each other along a vertical chain will have market power. For example, Shimano has considerable market power in the production of bicycle derailleurs and brakes, which it sells to Trek, Cannondale, Fuji, and other bicycle manufacturers. Likewise, United Technologies and General Electric have market power in the production of jet aircraft engines, which they sell to Boeing and Airbus, which in turn have market power in the market for commercial aircraft. How do firms along such a vertical chain exercise their market power, and how are prices and output affected? Would the firms — and would consumers — benefit from a vertical merger?

To answer these questions, we will consider the following example. Suppose an engine manufacturer has monopoly power in the market for engines. Suppose that an automobile manufacturer that buys these engines has monopoly power in the market for its cars. Leaving aside the costs and benefits of vertical integration discussed above, would this market power cause these two firms to benefit in any way if they were to merge? Would consumers of the final product — automobiles — be better off or worse of if the two companies merged? Many people would answer "maybe" to the first question, and "worse off" to the second question. People often raise objections to vertical mergers on the grounds that consumers will somehow be hurt. It turns out, however, that when there is monopoly power of this sort, a vertical merger is beneficial to the two firms, *and is also beneficial to consumers*.

To see that this is the case, consider the following simple example. Suppose a monopolist producer of specialty engines can produce those engines at a constant marginal cost c_E, and sells the engines at a price P_E. The engines are bought by a monopolist producer of sports cars, who sells the cars at the price P. Demand for the cars is given by

$$Q = A - P \,, \tag{1}$$

with $A > c_E$. To keep this example as simple as possible, we will assume that the automobile manufacturer has no additional costs other than the cost of the engine. (As an exercise, you can repeat this example assuming that there is an additional constant marginal cost c_A to produce the cars.)

First, suppose the two companies are independent of each other. The automobile manufacturer then takes the price of engines as given, and chooses a price for its cars to maximize its profits:

$$\max_{P} \ \Pi_A = (P - P_E)(A - P). \tag{2}$$

You can check that given P_E, the profit maximizing price of cars is given by:

$$P^* = \tfrac{1}{2}(A + P_E) \ , \tag{3}$$

and the number of cars sold and automobile company's profits are given by:

$$Q^* = \tfrac{1}{2}(A - P_E) \ , \qquad\qquad \Pi_A = \tfrac{1}{4}(A - P_E)^2 \tag{4}$$

What about the engine manufacturer? It chooses the price of engines, P_E, to maximize its profits:

$$\begin{aligned} \max_{P_E} \ \Pi_E &= (P_E - c_E)Q(P_E) \\ &= (P_E - c_E)\tfrac{1}{2}(A - P_E) \, . \end{aligned} \tag{5}$$

You should be able to easily confirm that the profit maximizing price of engines is given by:

$$P_E^* = \tfrac{1}{2}(A + c_E) \, . \tag{6}$$

The profits to the engine manufacturer are then equal to:

$$\Pi_E = \tfrac{1}{8}(A - c_E)^2 \tag{7}$$

Now go back to the expression for the profit to the automobile manufacturer, and substitute in the equation above for the price of engines. You will see that the automobile company's profit is given by:

$$\Pi_A = \tfrac{1}{16}(A - c_E)^2 \tag{8}$$

Hence, the total profits for the two companies are given by:

$$\Pi_{TOT} = \Pi_A + \Pi_E = \tfrac{3}{16}(A - c_E)^2 \tag{9}$$

Also, the price of cars paid by consumers is given by:

$$P^* = \frac{3A + c_E}{4} \ . \tag{10}$$

Vertical Integration. Now suppose that the engine company and the automobile company merged to form a vertically integrated firm. The management of this firm would choose a price of automobiles to maximize the firm's profit:

$$\max_P \ \Pi = (P - c_E)(A - P). \tag{11}$$

The profit-maximizing price of cars is now given by:

$$P^* = \frac{A + c_E}{2} \ , \tag{12}$$

and this yields a profit of:

$$\Pi = \tfrac{1}{4}(A - c_E)^2 \tag{13}$$

Observe that the profit for the integrated firm is *greater* than the total profit for the two individual firms that operate independently. Furthermore, the price to consumers for automobiles is *lower*. (To confirm that this is indeed the case, remember that $A > c_E$.) Hence, in this case vertical integration is of benefit not only to the merging firms, but also to consumers.

Why is this the case? The reason is that vertical integration avoids the problem of *double marginalization*. When the two firms operate independently, each one exercises its monopoly power by pushing its price above marginal cost. But to do this, each firm must contract its output. The engine producer contracts its output to push its price above its marginal cost, and then the automobile manufacturer does likewise. This "double marginalization" pushes the price above the "single marginalization" price of the integrated firm.

This example of double marginalization is illustrated graphically in Figure 7. The figure shows the demand curve (average revenue curve) for cars, and the corresponding marginal revenue curve. For the automobile company, the marginal revenue curve for cars is the demand curve for engines (effectively, the net marginal revenue for engines). It describes the number of engines that the auto maker will buy as a function of price. From the point

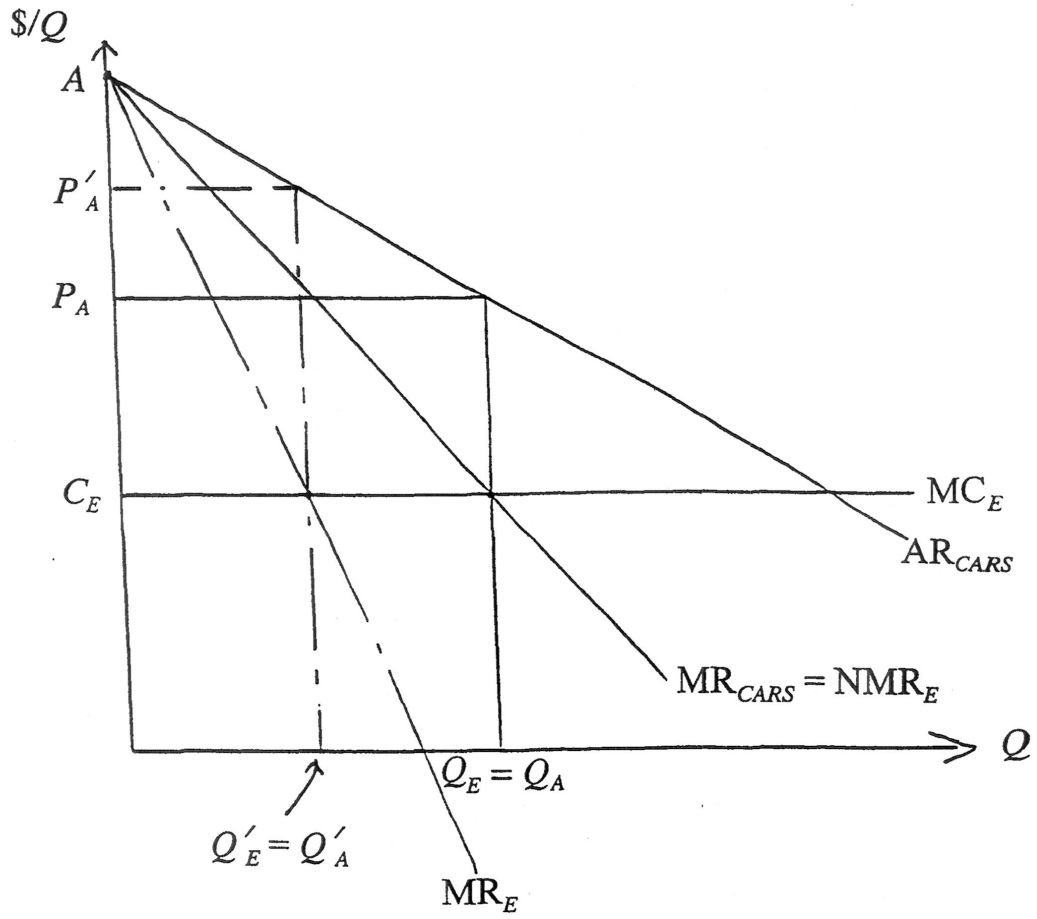

Figure 1: Example of Double Marginalization

of view of the engine company, it is the average revenue curve for engines (i.e., the demand curve for engines that the engine company faces). Corresponding to that demand curve is the engine company's marginal revenue curve for engines, labeled MR_E in the figure. If the engine company and automobile company are separate entities, the engine company will produce a quantity of engines at the point where its marginal revenue curve intersects its marginal cost curve. That quantity of engines is labeled $Q_E^{'}$. The automobile maker will buy those engines and produce an equal number of cars. Hence, the price of cars will be $P_A^{'}$.

What happens if the two companies merge? Then, the integrated company has the demand curve AR_{CARS} for cars and the corresponding marginal revenue curve MR_{CARS}. It produces a number of engines and equal number of cars at the point where the marginal

revenue curve for cars intersects the marginal cost of producing cars, which in this example is simply the marginal cost of engines. As shown in the figure, we then have a larger quantity of engines and cars produced, and a correspondingly lower price.

Alternatives to Vertical Integration. What can firms do to reduce the problem of double marginalization, assuming that a vertical merger is not an option? One solution is for the upstream firm to try to make the downstream market as competitive as possible, thereby reducing any double marginalization. Thus, Intel would like to do everything in its power to make sure that the market for personal computers remains highly competitive, and might even help firms that are in danger of going out of business. (Can you think of ways it could do this?)

A second method of dealing with double marginalization is called *quantity forcing*. The idea here is to impose a sales quota or other restriction on downstream firms so that they cannot reduce their output in an attempt to marginalize. We will discuss quantity forcing in more detail later in these notes when we turn to downstream distributers.

2 Monopsonistic Price Discrimination

In 15.010 you studied various forms of price discrimination by a firm with monopoly power. In the case of third-degree price discrimination, for example, you saw that when a firm can segment the market, it will charge a higher price to the group of consumers with the smaller elasticity of demand. You saw that when there are two groups of consumers, the firm should set prices and output levels so that the marginal revenue for each group is equal and is equal to marginal cost, i.e., $\text{MR}_1 = \text{MR}_2 = \text{MC}$. This leads to the following relationship that must hold for the prices:

$$\frac{P_1}{P_2} = \frac{1 + 1/E_2}{1 + 1/E_1}, \tag{14}$$

where E_1 and E_2 are the elasticities of demand for the two segments. (Note that this is Equation 11.2 in Chapter 11 of Pindyck & Rubinfeld, *Microeconomics*.)

A firm with monopsony power can also price discriminate, at least in principle. For

example, General Motors has considerable monopsony power in its purchases of automobile parts from various suppliers. General Motors can and often does exercise this monopsony power by "squeezing" suppliers with regard to output and delivery schedules, specifications, and prices. But for some of the parts it purchases, General Motors has more monopsony power than for other parts. In those cases where it has more monopsony power, it will tend to squeeze suppliers more. In effect, General Motors is practicing a very crude form of imperfect first-degree price discrimination.

We will begin by examining the analytics of third-degree monopsonistic price discrimination. We will see that this is exactly analogous to third-degree price discrimination by a firm with monopoly power. Next, we will examine *spatial* price discrimination by a firm with monopsony power – allocating purchases across regions to take advantage of differences in elasticities of supply. Of particular interest will be the market for timber, where paper mills, which have considerable monopsony power in buying timber, use regional allocations of purchases to price discriminate.

2.1 Third-Degree Monopsonistic Price Discrimination

Suppose a firm with monopsony power can segment the market into two groups with different price elasticities of supply. How much should it buy from each group, and what prices should it pay? To answer this, we use the same basic principles that apply to monopolistic price discrimination.

First, we know that however much the firm buys in total, its purchases should be divided between the two sources of supply *so that the marginal expenditure from each source is equal*. If this were not the case, the firm would not be minimizing its cost of purchasing. For example, if the marginal expenditure from the first group of sellers, ME_1, exceeded the marginal expenditure from the second group, ME_2, the firm could clearly do better by shifting its purchases from the first group to the second group. By shifting purchases this way, it would lower the price it paid to the first group (by buying less) and raise the price to the second group (by buying more). Hence, whatever prices the firm pays to the two groups, they must be such that the marginal expenditures for the two groups are equal.

Second, we know that *total* purchases must be such that the marginal expenditure from each group is equal to the marginal value of the product to the purchasing firm. If this were not the case, the firm would be better off either raising or lowering its total purchases (and hence raising or lowering its prices to both groups).

We can demonstrate these two principles algebraically. The *net value* of a purchase, NV, is the value of the purchase less the expenditures. Since the expenditure is coming from two sources, we can write this as

$$\text{Net Value } = \text{NV} = V - P_1 Q_1 - P_2 Q_2 \tag{15}$$

We want to maximize this with respect to Q_1 and Q_2. Differentiating with respect to Q_1 and setting the derivative equal to 0 gives

$$\frac{\partial \text{NV}}{\partial Q_1} = \text{MV} - \text{ME}_1 = 0 \tag{16}$$

Hence, we see that

$$\text{MV} = \text{ME}_1 \tag{17}$$

Likewise, for the purchases from the second group:

$$\text{MV} = \text{ME}_2 \tag{18}$$

Putting these together, we see that prices and purchases must be set so that:

$$\boxed{\text{ME}_1 = \text{ME}_2 = \text{MV}} \tag{19}$$

Again, marginal expenditure must be equal across groups of suppliers, and must equal marginal value. This is a key result, and as we will see, it will prove very useful when we examine spatial price discrimination in the next section of these notes.

We can translate this result into a rule for the relative prices that is analogous to eqn. (14) for monopolistic price discrimination. We can write marginal expenditure as:

$$\begin{aligned} \text{ME} &= P + Q\frac{dP}{dQ} \\ &= P\left[1 + \frac{Q}{P}\frac{dP}{dQ}\right] \\ &= P\left[1 + 1/E^S\right] \end{aligned} \tag{20}$$

where E^S is the elasticity of supply. Since this marginal expenditure must be the same for each group of suppliers, we have the following equation for the ratio of the prices:

$$\frac{P_1}{P_2} = \frac{(1 + 1/E_2^S)}{(1 + 1/E_1^S)} \tag{21}$$

Note that this is exactly the same equation that we had for third-degree monopolistic price discrimination. The only difference is that here the relevant elasticities are with respect to *supply*, not demand.

Figure 2 illustrates this for the case in which the marginal value (demand) curve is horizontal. (This corresponds to the case in which the firm is willing to buy any amount of the good, but only at a single price.) There are two groups of suppliers, and the relevant supply and marginal expenditure curves are shown. You can see that supply curve S_1 is less elastic than supply curve S_2. The firm buys quantities Q_1 and Q_2 from the two firms at the points where the marginal value curve intersects the marginal expenditure curve. Note that the firm buys at a *lower* price from the *less elastic* group of suppliers. Intuitively, the purchasing firm is taking advantage of the fact that the less elastic group is willing to supply significant quantities even at low prices.

Usually the marginal value (demand) curve will be downward sloping. Figure 3 illustrates this. In this case it is useful to horizontally sum the two marginal expenditure curves, obtaining a "total marginal expenditure" curve, ME_T. The total quantity purchased is found at the point where this total marginal expenditure intersects the marginal value curve. Since MV must equal ME_1 and ME_2, we can draw a horizontal line leftward from this intersection to find the quantities Q_1 and Q_2. The corresponding prices are found from the supply curves, S_1 and S_2. (If you have trouble understanding this figure, just compare it to Figure 11.5 in Pindyck & Rubinfeld; the two diagrams are almost exactly analogous.) Observe that in Figure 3, as in Figure 2, the group of suppliers with the less elastic supply curve receives the lower price.

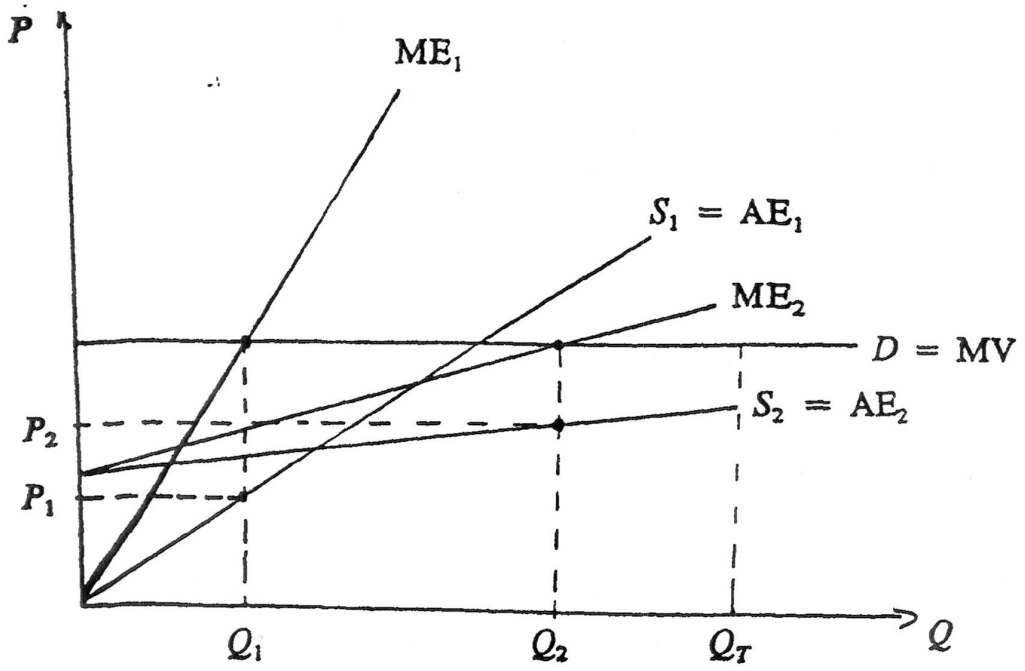

Figure 2: Monopsonistic Price Discrimination with Horizontal Demand

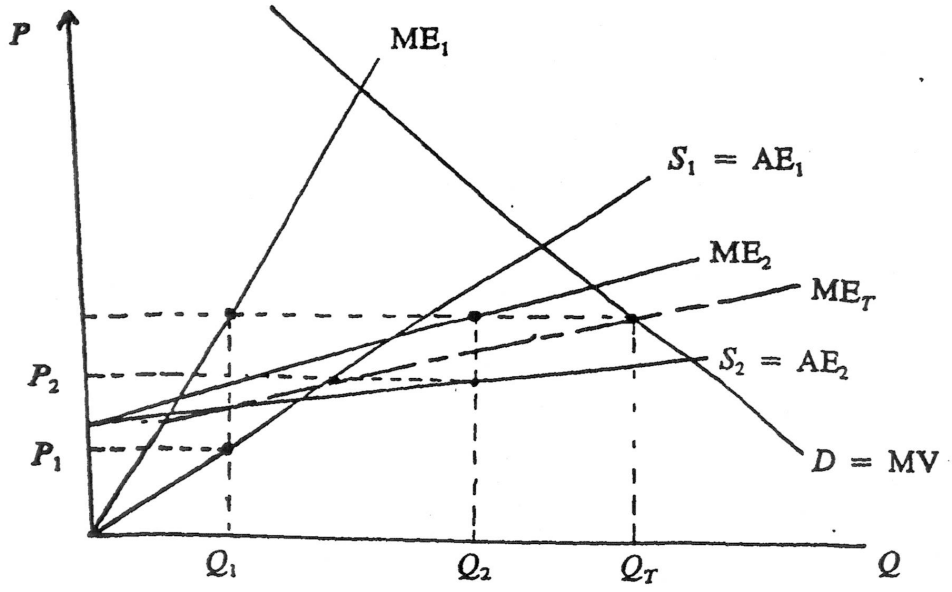

Figure 3: Monopsonistic Price Discrimination — General Case

2.2 Spatial Price Discrimination by a Monopsonist

There are occasions in which a firm with monopsony power can exercise that power via the regional allocations of its purchases. An example of this occurs in *timber markets* — in the United States and throughout the world. An analysis of monopsony power in this context arose as a result of the proposed merger of Georgia Pacific and the Great Northern Nekoosa Company in 1992.

When Georgia Pacific decided to acquire the Great Northern Nekoosa Company, several antitrust issues arose. Of particular concern was whether the merger would increase *monopsony power in timber markets*. At issue was whether the merged entity could depress timber prices, to the detriment of landowners.

Even after the merger there would still be a fair number of companies buying timber in the United States, so why the concern about monopsony power? The reason is that the costs of transporting timber to a mill are very high. As a result, markets for timber are *regionally localized*. (Can you think of examples of other markets that have this characteristic?) For example, a typical price for timber (at the site where it is cut) is about $10 per cord, and the cost of transporting it is about 20 cents per cord-mile. Hence the *delivered* price for timber from a source 100 miles from the mill can be two or three times as great as timber from a source only a few miles from the mill. As a result, mills usually consume timber from sources no more than 100 or 200 miles away, and often even closer. This means that mills that are, say, 300 or 400 miles apart do not compete with each other as buyers of timber.

The merger of Georgia Pacific and Great Northern Nekoosa created a situation where in some regions of Mississippi and Arkansas only one company would be purchasing timber, instead of two competing companies. The merged firm could then purchase timber at lower prices; indeed, from the point of view of Georgia Pacific's stockholders, this might have been seen as a benefit of the merger.

Should monopsony power be a concern here? Not at all, claimed Georgia Pacific. The company argued that even if the merger did result in only a single buyer of wood in some regional markets, that buyer would not be able to exercise any monopsony power. How could

that be? Their argument seemed quite simple. As you will learn if you end up working in the pulp and paper industry, *paper mills must always run at full capacity*. It is extremely costly to shut them down even temporarily, or to change their rate of output. Hence a paper mill's demand curve for timber is almost perfectly inelastic. This means that the quantity of wood the mill will buy must be the same whether or not it is a monopsony, and it would appear that it is unable to exercise monopsony power.

This is illustrated in Figures 4 and 5. Figure 4 is the standard diagram that shows how a firm exercises monopsony power. It describes a typical situation where the firm's demand curve for a factor input (i.e., its marginal revenue product curve) is downward sloping. The firm buys a quantity, Q_m at the point where its demand curve intersects the marginal expenditure curve. That quantity is less than the quantity that would be bought in a competitive market, Q_c, so that the price paid is lower. Figure 5 illustrates the situation for a paper mill, which has no choice but to buy a fixed quantity of wood. Because its demand curve for wood is vertical, it apparently cannot influence the price it pays for wood.

Or can it? We will see how the presence of high transportation costs can make it possible for a paper mill to exert monopsony power, even though it cannot alter the total quantity of wood that it purchases each month.

We will use a simple example. Suppose that initially a paper mill is buying timber from two locations, A and B. (This is illustrated in Figure 6.) Location A is only five miles from the mill, but Location B is 50 miles away. The cost of transporting wood is 20 cents per cord-mile, so the transportation cost from Location A is one dollar per cord, and from Location B is \$10 per cord. Suppose that the mill is initially buying 1,000 cords of wood from each location. Further assume that the competitive market price in Location A is \$15 per cord, and the competitive market price in Location B is six dollars per cord. (In other words, these are the prices that would clear the market if perfectly competitive buyers bought a total of 1,000 cords in each location.) Consistent with the competitive market, the mill is indifferent as to how much wood it gets from each location; the *delivered* cost of wood is \$16 in both cases. Thus the total cost of wood for the mill is $(2000)(\$16) = \$32,000$.

We will assume that the supply curve for wood in each location is isoelastic, and that the

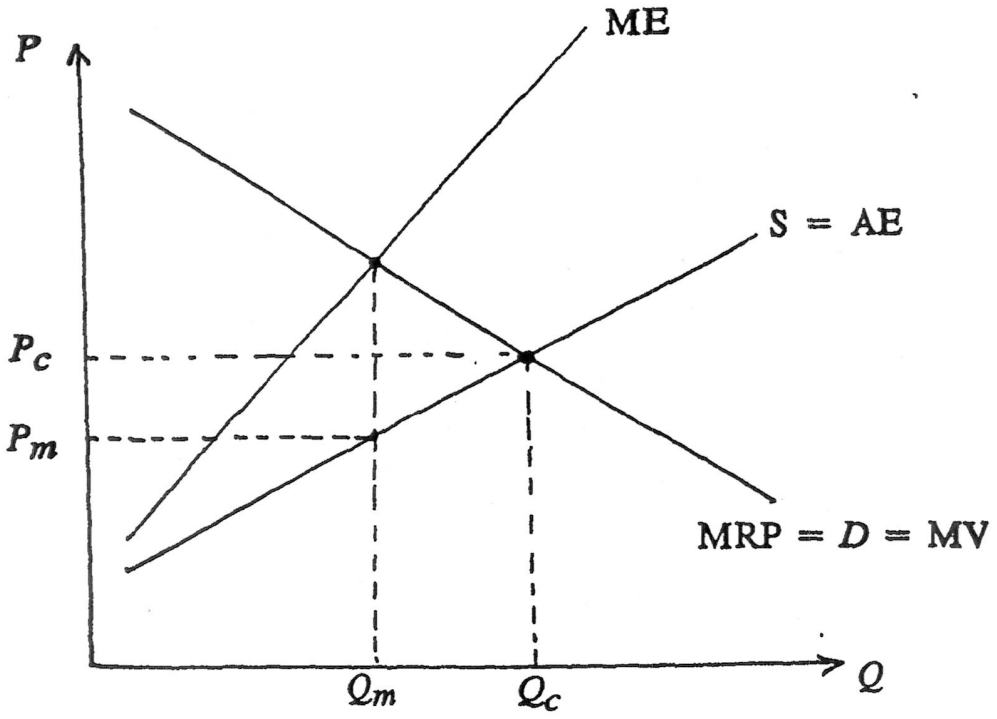

Figure 4: Monopsony Power

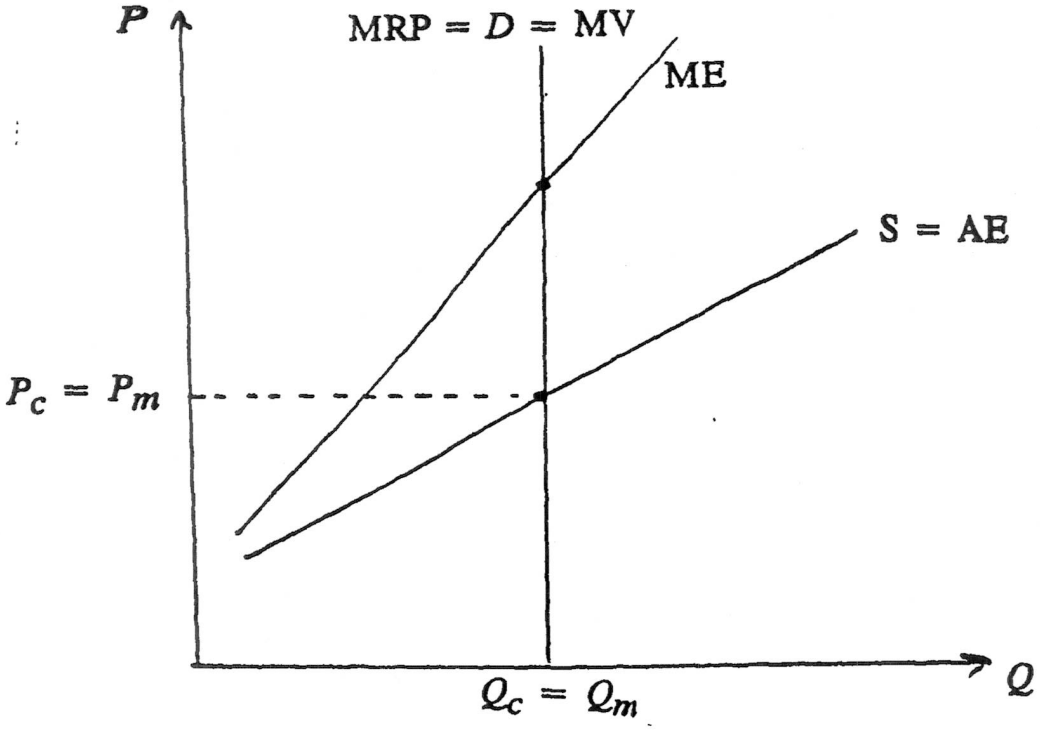

Figure 5: Monopsonist with Completely Inelastic Demand

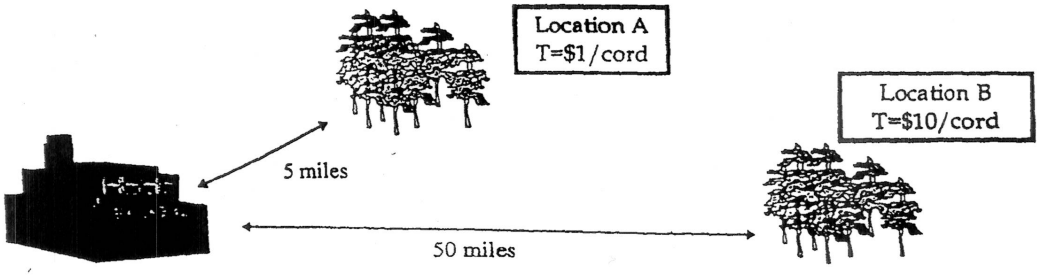

Figure 6: Wood Purchases by a Paper Mill

elasticity of supply is 0.5 in each location. You should be able to confirm that the competitive price and quantity of \$15 per cord and 1,000 cords, respectively, in Location A implies that the supply curve for Location A is given by:

$$\log Q_A = 5.554 + 0.5 \log P_A \tag{22}$$

To derive this supply curve, remember that a log-log demand or supply equation is isoelastic, and the price elasticity is given by the coefficient on the log of price. Since the elasticity is 0.5, we know the equation has the form $\log Q_A = a + 0.5 \log P_A$. To find the value of a, plug in the value of 1,000 for Q_A, and \$15 for P_A, and solve for a.

By going through the same steps, you should be able to confirm that the price and quantity of \$6.00 per cord and 1,000 cords in Location B implies that the supply curve for Location B is:

$$\log Q_B = 6.012 + 0.5 \log P_B \tag{23}$$

Now we will recognize the fact that the mill is a monopsonist, and see what happens if it reallocates its wood purchases between the two locations. Before, the mill was buying 1,000 cords of wood from each location, for a total cost of \$32,000. The mill still needs a total of 2,000 cords, but suppose it buys only 900 cords (10 percent less) from Location A, and 1,100 cords (10 percent more) from Location B. Since the price elasticity of supply is 0.5 in each location, this will cause the price in Location A to *fall* by approximately 20 percent, and it will cause the price in Location B to *rise* by approximately 20 percent. Hence the market price of wood in Location A would fall from \$15 to about \$12 per cord, and the

price in Location B will rise from \$6.00 to about \$7.20 per cord.[1] The mill must also pay a transportation cost of 1 dollar per cord from Location A, and \$10 per cord from Location B. Hence the *delivered* prices from these locations will be about \$13 and \$17.20, respectively.

What does this reallocation do to the mill's cost of wood acquisition? That cost is now given by:

$$\text{Cost } = (900)(\$13) + (1100)(\$17.20) = \$30,620 \tag{24}$$

Thus the cost of wood acquisition has fallen from \$32,000 to about \$30,620. The mill has indeed been able to exercise monopsony power and thereby reduce its wood acquisition costs.

To understand why this reallocation of purchases reduced the mill's acquisition cost, we can go back to the basic principles discussed in the previous section: marginal expenditure must be equal across groups of suppliers, and must equal marginal value. Let us calculate the mill's marginal expenditures when it was buying 1,000 cords from each location. Total expenditure on wood from any particular source is given by:

$$\text{TE} = P(Q) \cdot Q + T \cdot Q \tag{25}$$

where T is the transportation cost. Hence marginal expenditure is given by:

$$\text{ME} = (P + T) + \frac{\Delta P}{\Delta Q} \cdot Q \tag{26}$$

Remember that the elasticity of supply for each location is defined by:

$$E^S = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}\,, \tag{27}$$

so that

$$\frac{\Delta P}{\Delta Q} = \frac{P}{Q} \cdot \frac{1}{E^S} \tag{28}$$

[1] This is approximate because the 0.5 elasticity applies to a point on the supply curve, and we are looking at a movement along the curve. To get the exact prices:

$$P_A = \exp[(\log 900 - 5.554)/0.5] = \$12.14$$

$$P_B = \exp[(\log 1100 - 6.012)/0.5] = \$7.26$$

For a 10-percent quantity change, the approximation error is small, but for larger changes the error can be significant, so you should calculate the prices exactly.

Substituting this into our equation for marginal expenditure, we find that:

$$\text{ME} = (P + T) + \left(\frac{P}{Q} \cdot \frac{1}{E^S}\right) Q = (P + T) + \frac{P}{E^S} \tag{29}$$

If we now plug in the appropriate numbers for price and transportation cost, we find that marginal expenditure at each location is given by:

$$\text{ME}_A = (15 + 1) + 15/.5 = 46 \tag{30}$$

$$\text{ME}_B = (6 + 10) + 6/.5 = 28 \tag{31}$$

Marginal expenditure on wood purchased from Location A is much larger than marginal expenditure on wood from Location B. Hence we can do better by shifting some of our purchases away from Location A and to Location B.

In our numerical example, we examined what would happen if the mill bought 10 percent less wood from Location A and 10 percent more from Location B. There is nothing special about this 10-percent number, and it does not necessarily lead to cost minimization. (To see whether it does lead to cost minimization, you can calculate the marginal expenditures when the mill buys 900 cords from Location A and 1,100 cords from Location B, and see whether these marginal expenditures are now equal.) To minimize cost, we know that two conditions must be met: the marginal expenditures must be equal, and the total quantity of wood purchased must be 2,000 cords. We can write these conditions as follows:

$$P_A(Q_A) + 1 + \frac{P_A(Q_A)}{.5} = P_B(Q_B) + 10 + \frac{P_B(Q_B)}{.5} \tag{32}$$

$$Q_A + Q_B = 2000 \tag{33}$$

We can solve these equations for the optimal quantities Q_A^* and Q_B^*.

As an exercise, you will have an opportunity to explore in more detail how a paper mill can allocate its wood purchases optimally, so as to minimize its wood acquisition costs. At this point, however, it should be clear to you how high transportation costs provide another means for a firm to exercise monopsony power.

3 Downstream Distributers

Often, manufacturers rely on downstream distributers for the ultimate sale of their products. In some cases, these downstream distributers could have market power. As we saw before, this creates a problem of double marginalization. Other problems, however, arise as well. For example, some distributers might *free ride* on the efforts of other distributers or on manufacturers. For example, they might not do their share in advertising or otherwise promoting the products that they sell. Finally, distributers might face a problem of *"hold-up"* by suppliers; distributers might not want to invest in equipment or make other sunk costs tied to specific suppliers because then they will become dependent on the decisions of those suppliers.

We have already discussed how manufacturers can deal with the problem of double marginalization through *quantity forcing*. Automobile companies, for example, will often impose sales quotas (either directly or indirectly through nonlinear rebates and other sales incentives) on dealerships. Such dealerships often have local (regional) monopolies, and without such restrictions or incentives, they would want to impose high mark-ups on cars, even though that might mean selling fewer units than is optimal from the point of view of the manufacturer.

What about the problem of free riding? Often distributers must devote substantial resources in order to sell a product, e.g., promote the product, provide well-trained purchasing and sales personnel, maintain quality, etc. Typically, however, the distributer will not receive the full benefit from this expenditure of money and effort; some of the benefits will accrue to other distributers, and some will accrue to the manufacturer. Distributers therefore have an incentive to reduce these expenditures, and thereby free ride on the efforts of competing distributors and the manufacturer.

Free riding can be a particularly severe problem when the reputation of the product is important. Consider a fast food chain such as McDonald's or Burger King. These chains live or die by consumers' perceptions of product quality. (Most consumers visit a McDonald's not for an unusual culinary experience, but because they know exactly what they are going to

get, and can reasonably expect that food poisoning is not part of the package.) Maintaining high quality, however, is costly; each outlet must make sure that the ingredients are fresh, and that sanitary precautions are utilized. Suppose a McDonald's outlet decides to free ride on the McDonald's name by scrimping on quality as a means of saving some money. The result could be devastating for McDonald's.

As we will discuss later, *franchising* will not necessarily solve this problem. The franchisee will still have an incentive to free ride by reducing quality in an attempt to save money. Instead, McDonald's must impose direct controls on quality. This can be done by requiring franchisees (as well as company-owned stores) to buy ingredients only from approved suppliers, monitoring health and sanitary standards with a threat that an outlet that violates those standards will lose its franchise, etc.

Another method of dealing with the free-rider problem is to impose *territorial exclusivity*. This is common in the case of soft drinks and beer distributers. As is discussed below, in the case of soft drinks, territorial exclusivity is combined with quality standards as a means of dealing with the free rider problem.

3.1 Free Rider Problems in Soft Drink Distribution

The market power of Coca Cola and other soft drink companies resides largely in their ownership of syrup formulations and their brand names. To preserve this value, it is essential to maintain high-quality standards with respect to bottled soda. For example, it is essential that every bottle or can of Coca Cola contain precisely the right mix of syrup and carbonated water (with the correct degree of carbonation), that cans and bottles not be bent or chipped, etc. Given that bottling is done by downstream entities, Coca Cola must find a way to ensure that these quality standards are met. This can be difficult because of the free rider problem: it costs money to maintain quality standards, and a bottler may not receive the full benefit of those expenditures.

In addition, it is important that Coca Cola's soft drinks be properly advertised and promoted. Although national advertising is very important, it is also important to advertise and promote the brand at the local level. Because bottlers operate at the local level, they

are typically in the best position to do this advertising. The problem is that, once again, the bottlers may not be able to receive the full benefits of their advertising expenditures, and thus will have an incentive to free ride and do too little advertising. This is particularly the case when there are two or more bottlers all serving the same geographic area. Then, each bottler will have an incentive to sit back and let the other bottlers that are serving the area do the advertising.

3.2 Territorial Exclusivity

Soft drink companies like Coca Cola deal with these problems by giving *territorial exclusivity* to franchised bottlers. This helps to deal with problems of advertising and promotion, as well as quality control. Here are some advantages of territorial exclusivity:

1. Because each bottler has an exclusive territory, the returns from advertising expenditures have less spillover to other bottlers, and thus each bottler will have the incentive to advertise at close to the optimal level.

2. Territorial exclusivity also helps with the maintenance of quality standards because there are now fewer bottlers to monitor, and there is less likelihood of quality cutting from intrabrand competition by several different bottlers. The terms of the franchise contract allow for forfeiture if quality standards are not met, and the bottler knows that it has a local monopoly that is too valuable to jeopardize.

3. Another advantage of territorial exclusivity is that it forces small, inefficient bottlers out of the market. There are significant scale economies in soft drink bottling, and territorial exclusivity helps to achieve those economies.

4. Still another advantage of creating local monopolies is that it allows the bottlers to practice market-separating price discrimination. Specifically, a bottler can sell at different prices to different types of customers. If there were several bottlers competing with each other, this would not be possible.

Of course the creation of local bottler monopolies also creates problems. The bottler will now have the ability to capture profits that could have gone to the syrup company. In addition, because both the syrup company and the bottler now have monopoly power, we would expect to see double marginalization. Thus, syrup companies must find ways of extracting rents from the bottlers, and diminishing or eliminating the quantity-reducing effects of double marginalization. The syrup producers achieve these goals as follows:

1. To capture bottler profits, a syrup producer imposes franchise fees and royalty payments, as well as charging for the syrup itself. The fixed franchise fee and the royalty percentage (which applies to the bottler's gross profits) act as a two-part tariff. That two-part tariff, combined with control over the syrup price, give the syrup producer the ability to extract much (but not all) of the monopoly rents that would otherwise have gone to the bottler.

2. A particular form of quantity forcing is used to ensure that bottlers exceed their profit-maximizing output levels. Specifically, the franchise agreement requires them to serve *all* would-be customers in the area. Left to itself, a bottler would reduce output by refusing to serve small customers for which average servicing cost is high. The franchise agreement prevents this.

3. Finally, bottlers are forced to do most — but not all — of the advertising. The franchise agreement may call for specific levels of advertising, and the bottler must agree to undertake the expenses that this entails. It is still in the interest of the syrup producer to pay for *some* of this local advertising, because where one local monopoly abuts another local monopoly, there will be spillover from advertising and some incentive to free ride. Depending on the size and geographical boundaries of the bottler, the syrup producer can decide on an level of optimal co-payment for advertising.

Territorial exclusivity for downstream distributers is used in other industries as well. One example is automobile dealerships, which often have exclusivity over specified geographical

areas. Once again, problems of rent sharing and double marginalization arise. Can you think of ways that automobile companies can and/or do deal with those problems?

4 Franchising

Franchising is an increasingly common form of vertical structure. It has been estimated that about one-third of all retail distribution in the U.S. takes place through franchised outlets. The basic idea behind franchising is well summarized by an advertisement that Arby's ran seeking new franchisees: "You build the business, we'll build the brand." The franchisor (Arby's, in this case) develops a recognized brand name and a standardized service or product that has a (hopefully good) reputation. The franchisee sets up and runs a local outlet, in which the franchisor's services or products are sold. The franchisee must make a sizable initial investment in the outlet (perhaps with financial assistance from the franchisor), and then makes payments to the franchisor via a two-part tariff: an annual franchise fee, as well as a percentage royalty on sales.[2]

This kind of "business format" franchising is the focus of this section. It is quite distinct from the "trade name" franchising that we discussed in the previous section, i.e., where a franchisee simply distributes or sells a product manufactured by the supplying company that holds the trade name: examples include car dealerships, gasoline stations, and soft drink and beer bottlers. Although many of the same issues apply (e.g., free riding), the "business format" franchisee has more control over the way it runs its outlet.

Table 1 shows company-specific data for a variety of different franchise chains. (In putting this table together, I made no attempt to be comprehensive or representative.) The chains are grouped into restaurants, hotels, retail stores, and services. In each case, the table shows the total number of units in the chain, the percentage of the units that are outside of the U.S., the percentage that are company owned (as opposed to franchised), the average royalty rate paid by the franchisee, the average annual fee paid by the franchisee, and the average

[2]For an excellent recent study of franchising, see Roger Blair and Francine Lafontaine, *The Economics of Franchising*, Cambridge University Press, 2005. For detailed data on franchising, including company-specific data on individual franchising companies, go to http://www.worldfranchising.com.

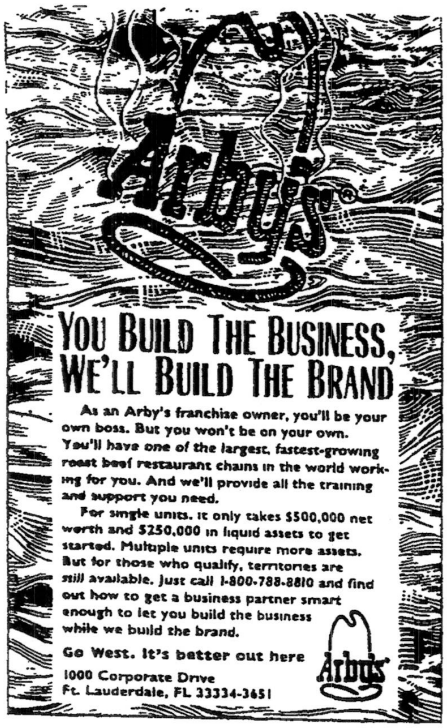

Figure 7: Arby's Advertisement

capital investment required to build and equip a new unit.

Why should a company like McDonald's franchise most of its restaurants, rather than own all of them outright? One reason is simply comparative advantage. McDonald's strength is its ability to develop new and appealing products, and develop and maintain a brand identity. The small scale entrepreneurs that become franchisees are often best able to organize and manage an individual local restaurant. Related to this, a franchisee is likely to have much better information about local demand patterns and local labor markets, and will be better able to hire and manage the employees in its outlet.

Nonetheless, not all McDonald's outlets are franchised: about one-quarter of them are owned and operated by the parent company. As Table 1 shows, other chains also have a mix of company-owned versus franchised outlets. How should McDonald's decide whether a new outlet should be franchised or company owned?

The answer largely has to do with the resolution of the free rider problem. If I own a McDonald's franchise but do a lousy job running it, I can still benefit from the McDonald's

Table 1: Examples of Chain Specific Data, 2009*

Chain	Years in Fran-chising	Total Number of Units	% Outside U.S.	% Company Owned	Royalty Rate (%)	Franchise Fee ($000)	Avg. Capital Required ($000)
A. Restaurants							
Arby's	44	3,704	3.3	31.4	4.0	37.5	1,418
Baskin Robbins**	60	4,700	49.2	0	5.9	40.0	337
Burger King	63	11,223	31.6	9.6	4.5	50.0	NA
Domino's Pizza	42	8,485	23.8	10.2	5.5	3.2	289
Dunkin Donuts**	53	7,200	29	0	5.9	50.0	955
KFC	56	5,310	NA	19.3	4.0	25.0	1,400
Long John Silver's	40	1,121	0	41.0	6.0	20.0	1,150
McDonald's	53	32,060	56.6	20.2	12.5	45.0	955
Subway	34	30,300	22.4	0	8.0	15.0	157
B. Hotels							
DoubleTree	20	154	2.6	49.7	4.0	75.0	2,010
Hampton Inn	25	1,412	2.1	2.8	5.0	50.0	6,800
Intercontinental	56	3,763	29.1	0.6	6.0	$500/Rm	3,000
Ramada	19	871	7.9	0	4.0	$350/Rm	5,290
C. Retail Stores							
7-Eleven	44	35,603	82.1	1.3	NA	64.0	118
Athlete's Foot	36	569	59.1	0	5.0	39.9	275
General Nutrition Centers (GNC)	18	5,781	35.2	64.8	6.0	40.0	157
Midas	42	2,591	39.9	3.6	10.0	20.0	305
Pearle Vision	28	952	14.6	55.4	7.0	30.0	375
D. Services							
Century 21 Real Estate	36	8,501	56.1	0	6.0	25.0	268
Coldwell Banker	27	3,508	17.2	24.9	6.0	19.0	336
Gymboree	30	536	45.7	0.5	6.0	45.0	150
Sylvan Learning Centers	28	1,119	8.0	17.7	8.5	46.0	225
Thrifty Car Rental	46	1,000	62.6	15.1	3.0	NA	225

* All data are from http://www.worldfranchising.com.

** Baskin Robbins and Dunkin Donuts are subsidiaries of Dunkin Brands, Inc.

name and the good job that *other* McDonald's outlets do. My incentive to free ride, however, is greatly reduced if most of my revenue comes from repeat business. Local customers know about – and care about – *your* outlet, not McDonald's in general.

Thus, McDonald's will franchise a new outlet if that outlet is likely to be a local restaurant, for which repeat business in critical. A McDonald's on the Connecticut turnpike or in the airport, on the other hand, is likely to be company owned, because there is little in the way of repeat business.

Now, here are some questions for you to think about.

1. It turns out that in franchised outlets, workers are paid lower wages than in company-owned outlets. Why should this be?

2. Note the differences in the two-part tariff for the chains listed in Table 1. For example, Midas has a relatively high royalty rate and relatively low annual franchise fee compared to the Athlete's Foot (a retail store selling running shoes and other sporting apparel). Does this make sense to you?

3. Ramada Hotels are all franchised, whereas DoubleTree owns about half of its hotels. Why might this be?